An Enterprise Geodatabase

Montgomery County, Maryland

Program Category #14: Information Technology

1.0 Abstract

The Montgomery County, Maryland Department of Technology Services - Geographic Information Systems (DTS-GIS) team has been providing, as one of the most important functions of the team, database services to County GIS users. The DTS-GIS data collection encompasses a wide range of spatial and non-spatial data including vector data, raster/imagery data and attribute tables such as Census 2000 publications. The County GIS data collection has been growing ever since 1992, the beginning of the County GIS program. It is now a terabyte inventory of data. Data originally required a single UNIX workstation for storage and has expanded over the years into a collection that occupies multiple UNIX and Windows workstations and mini-servers.

In 1996, DTS-GIS advanced its GIS database technology by creating its first geodatabase using Environmental Systems Research Institute (ESRI) Spatial Database Engine (SDE) technology. Since 1996, the geodatabase has been expanded, enhanced, and made more accessible to a greater variety of users including GIS specialists, IT specialists, engineers and desktop operators. The data is accessed through the use of GIS desktop software, internet and intranet applications. The County geodatabase, powered by the industrial standard database management system Oracle and coupled with the latest enhanced functionality of ESRI ArcSDE8.3, has improved significantly and has emerged from a small scale, GIS specific database into an enterprise level, multi-function database.

2.0 Need for the Program

Prior to the development of the centralized geodatabase, the County's DTS-GIS data inventory was stored on multiple UNIX and Windows data servers. In order to serve the data to users, a system administrator had to manage all disk drives, share them out to the public and mount or map the disk drives to users' desktops. For each County user interested in accessing the GIS data, the system administrator had to configure each individual desktop computer. This inevitably led to increased administrative responsibilities for the system administrator. As the GIS workforce in Montgomery County grew and demands for data access increased, the amount of administrative work grew proportionally. The County's search for a better solution led to the centralization of the County's geodatabase.

From an end-user perspective, there were two primary limitations of accessing the geo data from multiple disk drives rather than from a centralized database. First, for each disk drive to be mapped onto the user's PC, a unique drive is required. Since a Windows PC has at most 26 disk drives (A to Z), users often found themselves running out of available drives when trying to map additional GIS data disks. Second, for a user who utilized the GIS data infrequently, remembering the specific drive that hosted the data was a big challenge. Many times, users had to make rounds of telephone calls and spend hours of time to get a simple job done.

GIS web application developers who utilized the geo data in their applications were also required to memorize all data source locations. Often, when end users reported issues with the applications (such as a missing data layer), the developers had to spend a significant amount of time tracing back to troubleshoot the problem that related to data sources. The overhead for the web developers grew as the collection of GIS data grew.

3.0 Description of the Program

In 1996, DTS-GIS began the construction of a centralized geodatabase using ESRI's SDE technology. ArcSDE v8.0 was installed and configured on a Sun E4000 server equipped with two 167MHz processors. The database management system was Oracle v7.3.2. The deployment of the first geodatabase was successful; it hosted the property parcel data layer and all of the planimetric data layers – roads, building footprints, hydro features, transportation features, etc. Early ArcView users enjoyed the operation of simple connection and one-stop access of most of the spatial data layers.

Due to the fact that most of the GIS work was performed in a conventional Arc/INFO workstation environment before year 2000, and the fact that County GIS analysts mainly used spatial data in the form of Arc/INFO coverages, there was not a lot of effort devoted to enhancing the geodatabase. In 2001, the County geodatabase was upgraded to ArcSDE8.1 with database management system Oracle v8.1.6. With the release of ArcGIS desktop from ESRI in recent years, more GIS analysts and end users switched from Arc/INFO coverages and ArcView shapefiles to the use of feature classes in the ArcSDE geodatabase.

The year 2003 marked the turning point for the DTS-GIS geodatabase. By gaining the ownership of a more powerful UNIX server (Sun E450 with two 400MHz processors) with increased storage capacity, DTS-GIS was ready to transform the old geodatabase by migrating it from the old host machine into the new server. Time and labor were invested to make the new geodatabase a one-stop data service for all GIS analysts, web application developers and other end users. Taking advantage of increased functionality of ArcSDE8.3, the new geodatabase has been redesigned with a more logical structure. Data layers are organized into categories by natural relations. The new geodatabase is also able to serve a total of 45 gigabytes of color ortho-images as a single raster dataset; the original one foot resolution color ortho-image for the entire county was stored in 2672 TIFF files and spanned over 4 disk drives. By building eight layers of image pyramids, the geodatabase is able to serve the entire county image almost instantaneously. Many end users were amazed with the speed that this large dataset is retrieved and delivered from the database server to user's desktop. In addition, DTS-GIS is also working on storing all planimetric data layers in CAD drawing format into the geodatabase.

System and Architecture Layout

In 2003, the DTS-GIS data collection was consolidated into three major data servers. The latest addition (GIS-DB3) is the host of the County geodatabase. User workstations are connected to the data servers through the local Ethernet network at 100Mbps, which rides (at 1 giga-bit per second) on the County fiber-optic network backbone, as shown in Figure 1. The County's GIS program and use of technology have evolved over the years following the GIS industry leader ESRI's strategic development pattern. This is reflected in the County's GIS computing environment layout, which is shown in Figure 2.

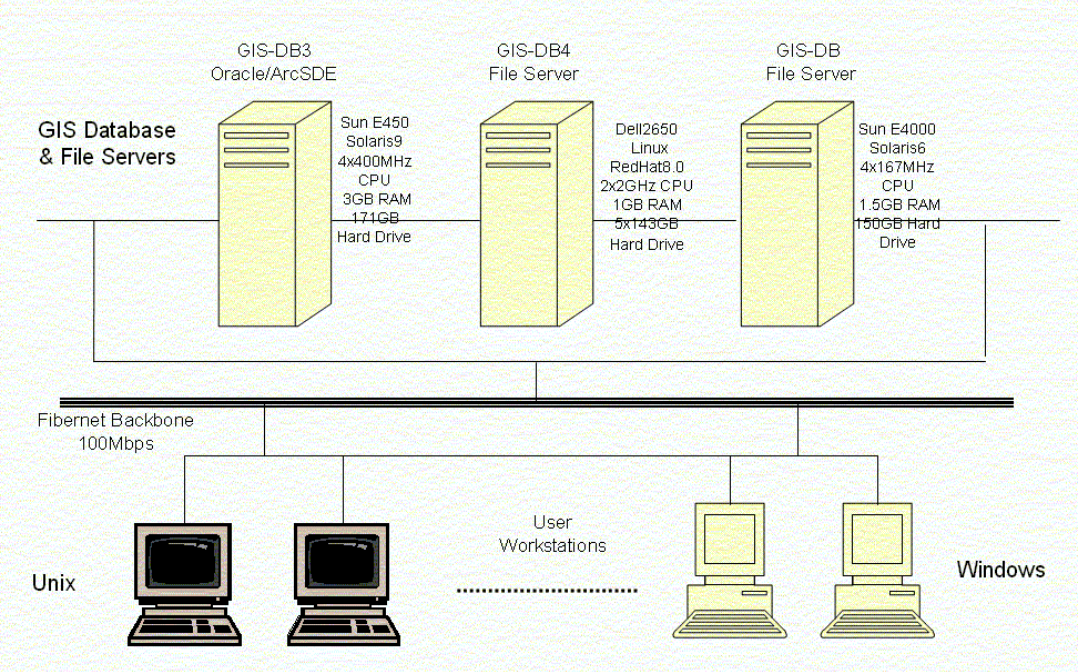

Figure 1. GIS Data Server System Architecture

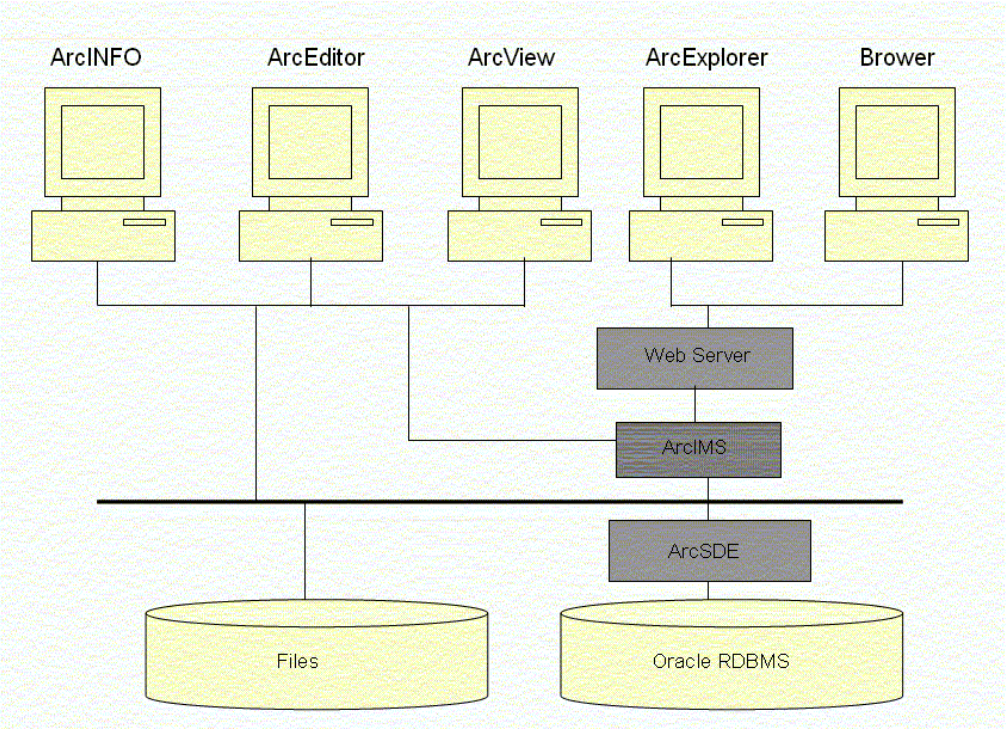

Figure 2. GIS Computing Environment

Database Software Installation, Configuration and Tuning

Although ArcSDE8.3 is compatible with both Oracle v8.1.7 and Oracle v9.0.4, the decision was made to install Oracle v8.1.7 because of its proven stability. Following the successful installation of the Oracle database management system, ArcSDE8.3 was installed. A considerable amount of time was spent

configuring and tuning the Oracle instance and the ArcSDE service because the database performance is completely dependent on the configuration and tuning of the Oracle and ArcSDE. Following both Oracle and ESRI configuration guidelines and utilizing the server's three gigabyte memory space, the DTS-GIS staff has successfully configured and tuned the database and hence optimized the performance.

Database Design and Data Organization

Montgomery County's GIS data collection comprises different data types, including vector, raster, CAD drawings and attribute tables. Good database design and data organization are essential to facilitating user orientation and simplifying accessibility of the data. Taking advantage of newly introduced geo-dataset concepts in ArcGIS and ArcSDE, vector data layers are organized into different datasets, each containing naturally associated data layers. For example, the DISTRICTS dataset contains all service/administrative area polygons, such as election districts, school service areas, fire boxes, police districts etc…; the LOCATIONS dataset contains all location points, such as points (or places) of interest, hospital locations, school locations etc... Figure 3 captures a snapshot of the content of the datasets NATURE&REC (nature and recreation); it features bikeways, parks, scenic routes and trails.

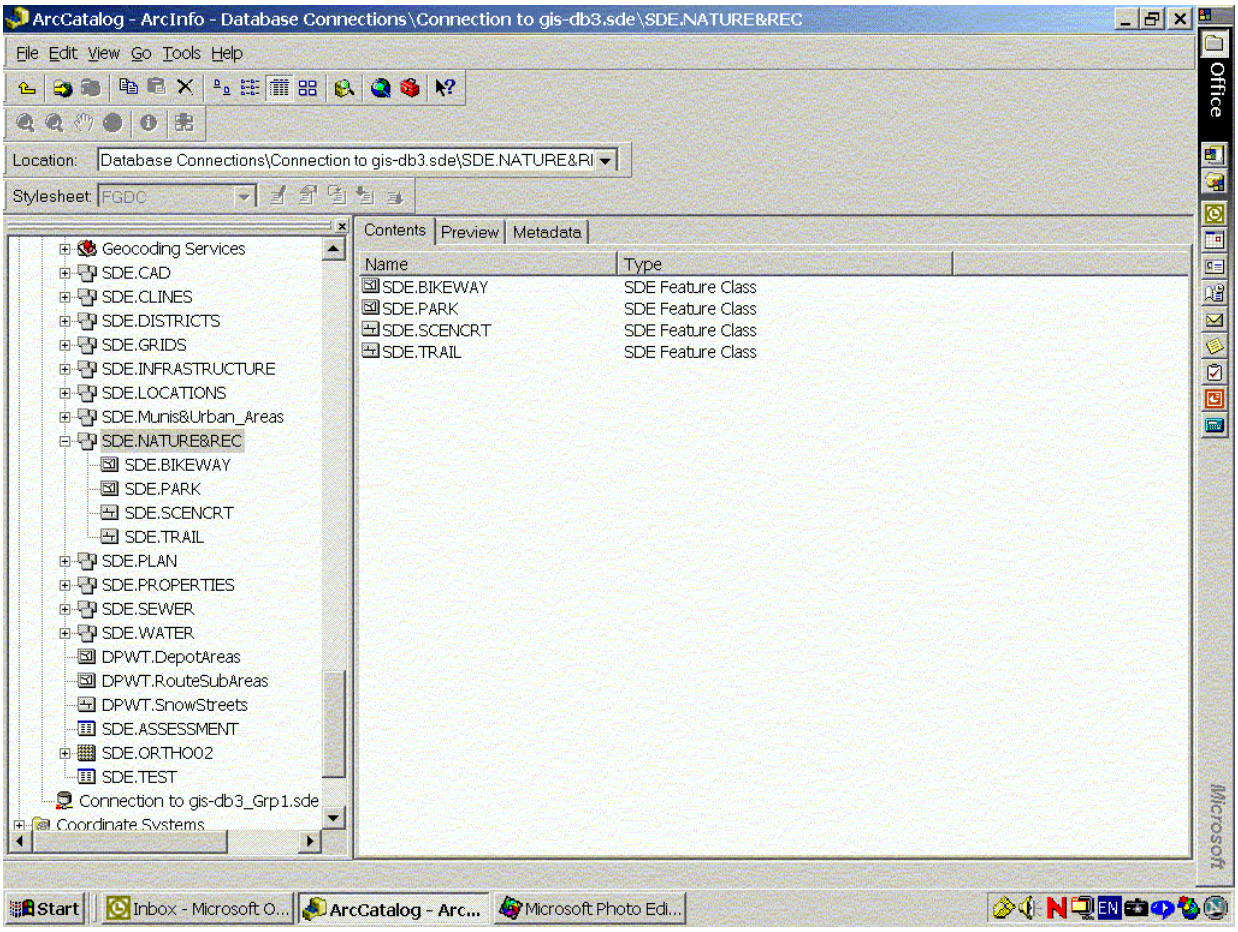

Figure 3. Content of Dataset NATURE&REC

CAD drawings are stored in a separate category; each planimetric data layer is stored as a feature class in the CAD dataset. The 2002 color ortho-image of the County is stored in a single raster dataset ORTHO02. Figure 4 shows the list of datasets in the DTS-GIS geodatabase.

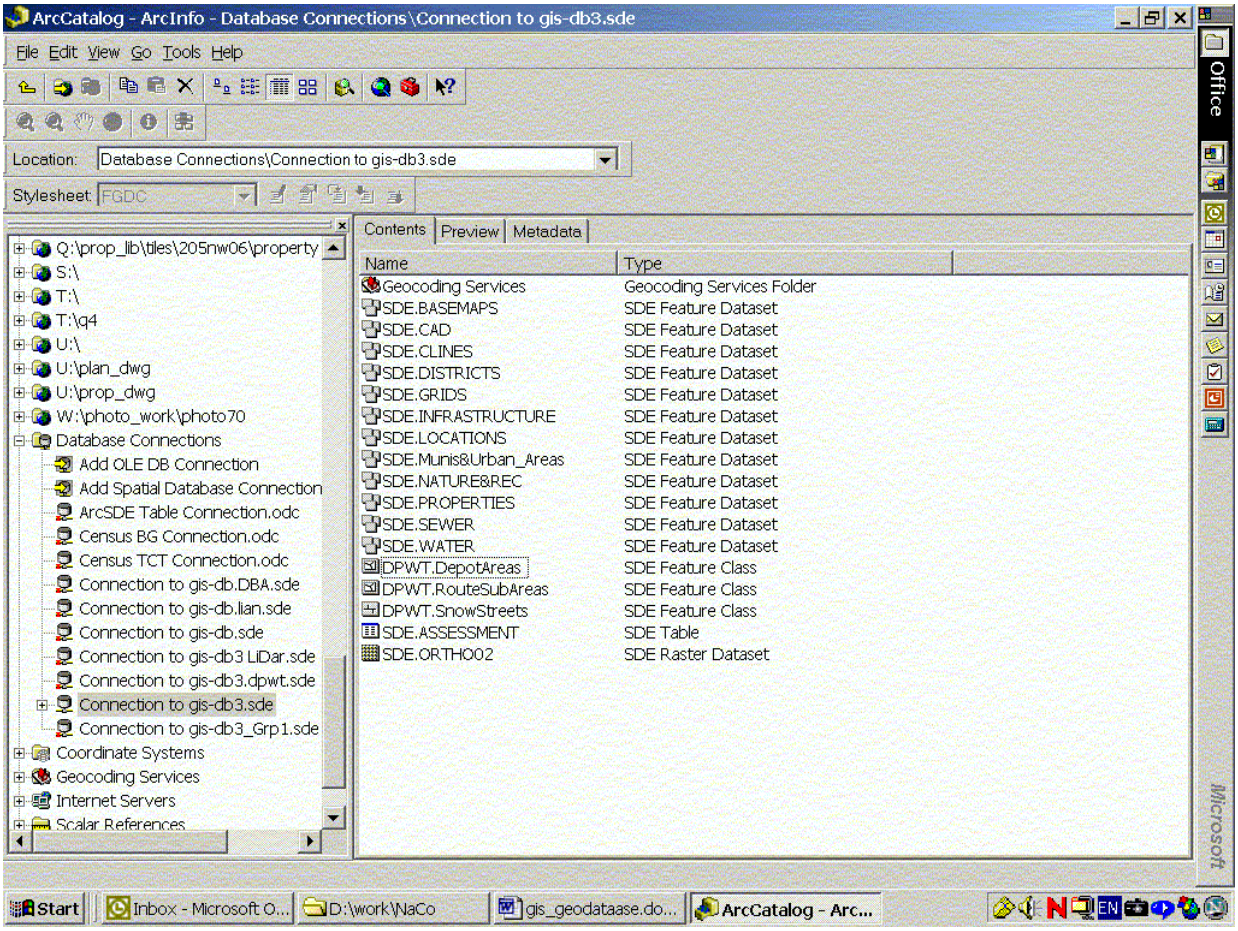

Figure 4. Datasets in DTS-GIS Geodatabase

Database Completion and Testing

Once the database was designed, the geo data was loaded into the centralized geodatabase from different data servers and disk drives. Vector data was loaded using the ArcGIS catalog manager ArcCatalog; raster data was loaded using ArcSDE command SDERASTER. Uploading the entire county raster images of 2672 TIFF files into a single ArcSDE raster dataset represented quite a challenge. The prerequisite for merging multiple image files into a single raster data map in ArcSDE is that each image file has to have the exact pixel registration as its neighboring image files, which means that pixels from multiple image files line up exactly, and the corners of those image files meet at the exact x, y coordinates. A batch job was written to clean up the TIFF files to ensure that all adjacent image files had the same coordinates. Another batch job was executed to merge all TIFF files into a single mosaic while the TIFF files were loaded into ArcSDE one by one. Eight layers of image pyramids (each has reduced resolution for the purpose of fast rendering of images) were built for the raster dataset. Finally, the dataset was analyzed using Oracle optimization tool. Test results revealed that the speed for raster image retrieval and rendering greatly outperformed the previous method of serving images through the use of individual TIFF files. Staff investment in this effort proved to be very worthwhile. The geodatabase was initially released to DTS-GIS analysts for testing. The process was generally smooth and most reports and feedback were positive.

Deployment of the Geodatabase

Since both Oracle and ArcSDE communicate to their client software through the TCP/IP protocol, deployment of the geodatabase to all users and departments is much easier than the old way of mounting file systems. Most end users were able to enter the database server information into their desktops by following a set of simple instructions. Few end users required phone services or office visits. The deployment was quick and smooth, yielding a big savings on the GIS system administrator's time.

4.0 Use of Technology

The Montgomery County geodatabase is built based on state of the art technology from the industry leaders Oracle and ESRI. By combining the power of Oracle database technology and ESRI's ArcSDE, DTS-GIS has created a scaleable, user friendly, centralized data warehouse. With open data access capability in mind, DTS-GIS staff also employed a suite of open database connection technologies such as Oracle and SQL/Server ODBC, Oracle and SQL/Server OLE DB to link the geodatabase to other County databases. These databases include the Division of Solid Waste Services' customer service database, the Department of Housing and Community Affairs housing development database and the centralized County web server database. The technologies have made the users' desire for "one-stop access to multiple data sources" come true.

5.0 The Cost of the Program

The migration of the geodatabase from the old host machines into the new server with upgraded software marked the beginning of the DTS-GIS enterprise geodatabase. It took a full time database administrator three months to complete the process including installation of the Oracle and ArcSDE services; configuration and tuning of the database; and database implementation and deployment. The total labor cost is estimated at $30,000. The Server machine costs $90,000 and the software (Oracle and ArcSDE) licenses cost $27,000 . The hardware and software costs add up to $117,000. One full time staff member is devoted to both database administration and system administration on the server side. System administration on the client side is minimal.

6.0 The Results/Success of the Program

The successful implementation and deployment of the County geodatabase dramatically improved the DTS-GIS data service quality. The advantages and improvements it brought to the County include:

1. Fast data access
2. Easy search for and retrieval of the desired data
3. Easy installation and configuration of client connections
4. Reduced system administration time
5. One stop access to multiple data types from multiple data sources

7.0 Worthiness of an Award

A well-designed and easy-to-use robust database is the foundation for successful IT applications. Montgomery County DTS-GIS is currently managing and serving a vast amount of spatial and non-spatial data via this state of the art, enterprise geodatabase. By providing data sharing and data integration

capabilities through database connections to multiple databases, the geodatabase is not only serving the GIS community but also IT specialists, engineers, desktop operators and web applications. With easy access to information, more and more end users and applications are tapping into this centralized data warehouse.

The fast dissemination of integrated information is also bringing substantial benefit to citizens. Through the use of web technology, citizens are obtaining their personal and neighborhood information from the database. For example, personal real property information, recycle/refuse days and times, neighborhood snow removal status and map services are all accessed utilizing data in the geodatabase. The return on investment has been substantial and will continue to increase.